

Fully Automatic and Real-Time Catheter Segmentation in X-Ray Fluoroscopy

Pierre Ambrosini^{1,3(✉)}, Daniel Ruijters², Wiro J. Niessen^{1,3,4},
Adriaan Moelker³, and Theo van Walsum^{1,3}

¹ Biomedical Imaging Group Rotterdam, Department of Medical Informatics,
Erasmus MC, Rotterdam, The Netherlands

p.ambrosini@erasmusmc.nl

² Philips Healthcare, Interventional X-ray Innovation, Best, The Netherlands

³ Department of Radiology and Nuclear Medicine, Erasmus MC,
Rotterdam, The Netherlands

⁴ Imaging Physics, Faculty of Applied Sciences, Delft University of Technology,
Delft, The Netherlands

Abstract. Augmenting X-ray imaging with 3D roadmap to improve guidance is a common strategy. Such approaches benefit from automated analysis of the X-ray images, such as the automatic detection and tracking of instruments. In this paper, we propose a real-time method to segment the catheter and guidewire in 2D X-ray fluoroscopic sequences. The method is based on deep convolutional neural networks. The network takes as input the current image and the three previous ones, and segments the catheter and guidewire in the current image. Subsequently, a centerline model of the catheter is constructed from the segmented image. A small set of annotated data combined with data augmentation is used to train the network. We trained the method on images from 182 X-ray sequences from 23 different interventions. On a testing set with images of 55 X-ray sequences from 5 other interventions, a median centerline distance error of 0.2 mm and a median tip distance error of 0.9 mm was obtained. The segmentation of the instruments in 2D X-ray sequences is performed in a real-time fully-automatic manner.

Keywords: Catheter · Guidewire · Tracking · X-ray · Fluoroscopy · Deep learning · Convolutional neural network · Segmentation

1 Introduction

Minimally invasive procedures are generally preferred over open surgery interventions, as these localized and accurate interventions lead to less trauma and shorter recovery times than conventional procedures. Minimally invasive procedures require real-time imaging to visualize the relevant anatomy and the instruments. Particularly, in catheterization procedures, a catheter¹ is inserted

¹ For clarity, in the remainder of the paper, the word “catheter” also refers to the micro-catheter and guidewire instruments. Although they have quite different appearances, they are handled altogether as one instrument in the proposed method.

into the body via the vasculature and fluoroscopic imaging is used to continuously visualize the catheter. The vasculature is only visible in X-ray images when contrast agent is injected, and contrast agent is used sparingly because of its toxic nature. Therefore, recent approaches for virtual roadmapping that permit the visualization of a 3D vessel tree from pre-operative images have been presented [1, 12]. Such methods benefit from automated extraction of the instruments from fluoroscopic images. The purpose of this work was therefore to develop and evaluate a method that segments fully automatically the catheter in 2D single-plane X-ray fluoroscopic sequences in real-time.

Automatic catheter segmentation is not straightforward, as the catheter is a thin, moving structure with low contrast in noisy images. Segmentation methods for electrophysiology (EP) electrodes and EP catheter in 2D X-ray images have been reported [3, 18]. EP electrodes are clearly visible, and their location is often used to obtain a full segmentation of EP catheters. Segmentation of catheters without features such as electrodes has been studied less frequently. Most methods enhance the instruments with Hessian-based filters, which are followed by a spline fitting approach, starting from the catheter shape of the previous frame [2, 4, 5, 7, 13, 17]. These methods have two drawbacks: the first frame of the fluoroscopic sequence has to be manually annotated and the curvature and length of the catheter should not change much between frames. [4, 7] propose semi-automatic methods to segment the first frame. Recently, a fully automatic method using directional noise reduction and path extraction, with segments and similarity from the previous frame cost function, has been proposed [16]. The method was evaluated on the last frame of 7 sequences from one canine study on which it performs well; it is, however, not a real-time method. We summarize in Table 1 the methods proposed in the literature in order to show results they obtained. Note that the results, the accuracy metrics and computation times cannot be directly compared but they give an idea of the performances.

Our method utilizes deep convolutional neural networks (CNNs) for the segmentation. CNNs have been demonstrated to be very effective in image classification and image segmentation [9], also in case of medical images with a

Table 1. Summary of the methods in the literature and the method of this paper.

References	Fully auto.	Time	Accuracy	Tip accuracy
2003 Baert et al. [2]	No	5 s	Mean 0.9 px ^a	Mean < 2 mm ^a
2007 Slabaugh et al. [13]	No	175 ms	-	-
2009 Wang et al. [17]	No	500 ms	Mean 2 px (0.4 mm)	Mean 5.4 px
2012 Heibel et al. [7]	No	60 ms	Mean 0.8-3.9 px	-
2016 Chang et al. [4]	No	-	-	-
2016 Chen et al. [5]	No	-	Mean 2.1 px (0.5 mm)	-
2016 Wagner et al. [16]	Yes	>1 min	Mean 0.5 mm	-
This work	Yes	75 ms	Median 0.2 mm	Median 0.9 mm

^aThe failed segmentations are not included in the evaluation

limited set of annotations [10, 11]. Ronnerberger et al. [11] introduced an end-to-end biomedical imaging segmentation network called U-net: a model with a fully convolutional part (downsampling part) and a deconvolutional part (upsampling part) which outputs after appropriate thresholding a binary segmented image. Extensive data augmentation enables the neural network to generalize well, even in case of small training sets. Several improvements w.r.t. the network and the training process have been introduced more recently. The downsampling (resp. upsampling) part has been shown to be more effective with strided convolution (resp. transposed convolution) than with max pooling [10, 14]. Strided convolution enables to learn how the features should be downsampled/upsampled. Moreover, batch normalization [8] and residual learning [6] have been proposed to improve training convergence.

2D X-ray fluoroscopic images are very noisy and are used in liver catheterization procedures to guide catheter inside the liver vessel tree. The catheter used does not have specific features, such as electrodes, nor specific shapes that may facilitate their segmentation. The framerate is around 7 Hz, so, during the catheter manoeuvre, its shape and length may change considerably between two consecutive frames. Tracking the catheter over time is therefore quite challenging. In this paper, we propose a fully automatic segmentation method based on the U-net model combined with recent strategies to improve the training of the network, such as batch normalization, residual learning and a data augmentation scheme to increase the size of the training dataset. The catheter centerline is then extracted using skeletonization and linking of the extracted branches. Our work differs from previous approaches in that we introduce a fully automatic approach that can be run in real-time.

2 Method

The catheter is segmented using the CNN and then the centerline is extracted from the result of the CNN using skeletonization and subsequent linking of branches.

2.1 Data

A 2D X-ray image sequence is a set of s 2D images $S = \{I_1, I_2, \dots, I_s\}$. Each image I_i is associated with a binary image B_i where the catheter pixels have a value of 1 and the background pixels 0. We also associate the output from the neural network, the image prediction B_i^p , where each pixel is between 0 and 1, and a pixel closer to 1 is considered as a catheter pixel. Our neural network model is trained to predict a mask B_i^p , given an image I_i as input.

2.2 CNN Model

The model is an adapted version of the well-known U-net model [11]. The input of the model is the current image I_i and previous frames I_{i-1}, \dots, I_{i-3} . The output

is the image prediction B_i^p . For the network topology, see Fig. 1. To improve convergence speed during training, we add batch normalization (BN) after every convolution [8]. In order to also learn how to downsample/upsample the features we use strided convolutions [10, 14]. To prevent overfitting we add dropout at the end of the two last blocks in the downsampling part [15].

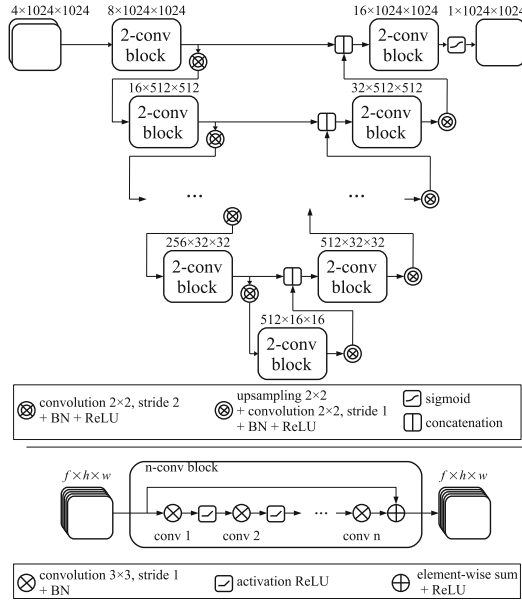


Fig. 1. Neural network model (top): $I_i, I_{i-1}, \dots, I_{i-3}$ images are fed into the model and segmented image B_i^p is predicted. The model is composed of n-conv block at each layer. An n-conv block is n consecutive convolutions of input features (with f filters, height h and width w) with a residual connection to the output [6, 10] (bottom).

2.3 Training

The loss function is based on the Dice overlap metric [10] between the ground truth mask B_i and the output image B_i^p of the model, defined as:

$$L_{\text{Dice}}(B_i, B_i^p) = - \frac{2 \sum_k B_{ik} B_{ik}^p}{\sum_k (B_{ik}) + \sum_k (B_{ik}^p)} \quad (1)$$

where B_{ik} and B_{ik}^p are respectively the pixels of the mask B_i and the output image B_i^p .

In order to have more data to train and to generalize well, large data augmentation is used during the training. Data augmentation is done on the fly. For every image in the set of training images, there is a 50% probability to augment

the image. If the image is augmented, we apply all of the following transformations to both the X-ray image inputs and the corresponding binary image output: 50% probability on a horizontal and vertical flip, a random rotation (around the image center) in a range of $\pm 9^\circ$, a random scale with a factor in a range of 0.9 to 1.1, a random horizontal and vertical translation in a range of $\pm 16\%$ of the image size, a random intensity shift with a factor in a range of ± 0.07 (in normalized image between 0 and 1) and a Gaussian noise with $\sigma = 0.03$.

2.4 Centerline Extraction

The output of the neural network is first thresholded with a threshold α (between 0 and 1) and then skeletonized [19]. Next, the branches (ordered sets of pixels) are determined based on connectivity. Connection points are created between close branches. If the closest points between two branches are within a distance D_{\max} pixels, we consider this to be a possible connection, and there can be only one connection between two particular branches. Then, to link the branches, three steps are done (Fig. 2). First, for each connection, we divide and merge branches in order to have the longest branches. Second, loops are detected and merged following the direction at the crossing point. We have a loop in a branch when two points, within a distance D_{\max} pixels, have their distance along the branch of at least B_{\min} pixels. Before the third step, the first and second steps are repeated a second time with a distance D'_{\max} superior to D_{\max} . Finally, in the last step, the remaining connected branches larger than B_{\min} pixels are considered as incomplete loops or straight loops due to foreshortening. We process them similarly as the second step by closing the two endpoints of their branch. When all the potential links have been processed, we keep the longest connected set of branches and choose amongst the two endpoints the farthest from the image border as the tip of the catheter centerline. Finally, the centerline is smoothed by fitting a spline.

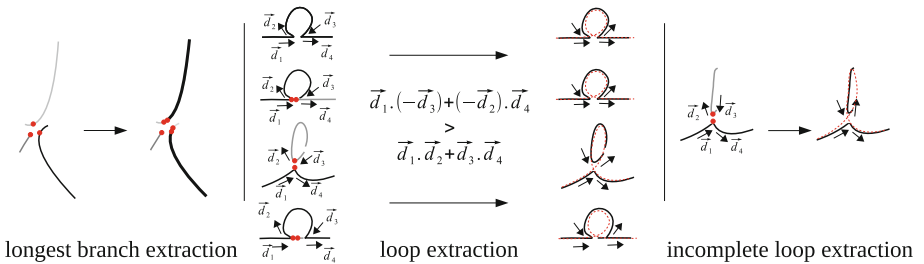


Fig. 2. The three steps for the branches linking. Branches are depicted with different colors and connection points are in red. The red dash curve is the centerline after spline smoothing. The first step links the longest branch parts, the second step links and corrects the loops and the last step links the remaining branches longer than B_{\min} , considered as an incomplete loop or a straight loop due to foreshortening.

3 Experiments and Results

2D single plane X-ray fluoroscopic sequences have been acquired during 28 liver catheterization procedures in three different hospitals (Erasmus MC, Rotterdam, the Netherlands; Hôpitaux Universitaires Henri Mondor, Créteil, Paris, France; and Ospedale di Circolo e Fondazione Macchi, Varese, Italy) with angiographic C-arm systems (Xper Allura, Philips Healthcare, Best, the Netherlands). Every image is normalized on the range $[0, 1]$, mapping the intensities between the 2nd and the 98th percentile of the image histogram. In 182 sequences from 23 procedures, we manually segmented the catheter in four consecutive images by annotating points and fitting a spline. From the catheter spline we constructed the binary segmented image using a dilation operator with a 5×5 pixel kernel. These sequences were used as training data for the parameter optimization. In 55 sequences from the other 5 procedures, we also segmented four frames per sequence. These 55 sequences will be used as testing data after the model optimization and training.

The loss function L_{Dice} of the model is optimized using stochastic gradient descent with a learning rate of 0.01, a decay of 5.10^{-4} and a momentum of 0.99. Following the training, we set the threshold α to 0.01, the maximum distance to connect two branches during the first pass (resp. second pass) D_{max} to 5 pixels (resp. D'_{max} to 20 pixels) and the minimum loop length B_{min} to 30 pixels. Using an Nvidia GTX 1080, the average time to segment one image was 75 ms which is suitable for real-time processing. Our method is publicly available².

We evaluate using the tip distance error (i.e. the distance between the annotated catheter tip and the tip of the segmented catheter), and the average distance between the manually segmented catheter and the automatically segmented catheter. Figure 3 shows the tip and catheter distances results. We compute the precision of the tip between consecutive frames. The median, average, minimum and maximum of the standard deviation per sequence of the tip distance error are respectively 0.7 mm, 5.0 mm, 0.1 mm and 55.7 mm. Five examples of segmentation are shown in Fig. 4. In the third frame, the segmentation is going too far and follows part of the vertebrae. The fourth frame misses the proximal part of the catheter. The last frame is the only sequence with significant false positives. It is less noisy because it has been acquired with higher radiation dose. The neural network was not trained for such sequence.

4 Discussion and Conclusion

We proposed a fully automatic method to segment catheter on 2D X-ray fluoroscopic images using CNNs. The segmentation on testing data gives a median tip distance error of 0.9 mm and a median centerline distance error of 0.2 mm where 85% of the frames have less than 1 mm of centerline distance error. We note that the distance errors are in mm at the X-ray detector scale. The real distance errors at the patient scale are smaller.

² Available at <https://github.com/pambros/CNN-2D-X-Ray-Catheter-Detection>.

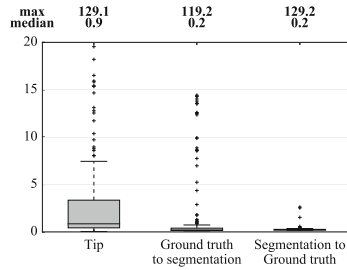


Fig. 3. Tip distance error, ground truth to segmented centerline distance error and segmented centerline to ground truth distance error (in mm) on the test set: 4 frames per sequence on 55 sequences.

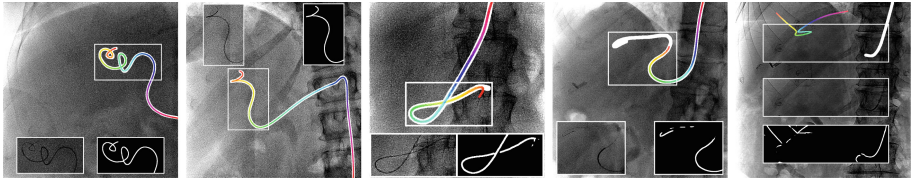


Fig. 4. 5 segmented frames of the test set. The segmented centerline (dilated for visual purpose) appears from red (tip) to rose. The ground truth (thicker) appears in white. The windows show the original image and the output of the neural network.

Very few images have false positives after the CNN segmentation. Therefore, we can use simple criteria to extract the catheter centerline from the CNN segmentation. The results show that it works well and can handle self-intersections. The main problem in the extracted catheters are sometimes large gaps in the segmentation due to false negatives. As a consequence, occasionally the proximal part of the catheter is missing. With a larger training set, the model is expected to generalize better.

Previous studies show a higher success rate, probably because they manually initialize the tracking process. We, in contrast, do not employ a tracking approach. Whereas it is clear that a stronger incorporation of the time dimension (beyond using consecutive frames in the segmentation) may provide a more robust result, our current results demonstrate that even without tracking good results can be obtained. A major advantage of not utilizing tracking is that the method is not hampered by previously incorrectly segmented frames, and thus automatically can recover from previous failures.

The catheter and guidewire have different thickness and appear quite differently on fluoroscopic images. Whereas we trained one network that segments both, it could be interesting to use two different models and retrospectively combine their results to obtain a more accurate segmentation. The current model is using the previous images to segment the catheter but it could also be useful to use the previous segmentation in the model. Both strategies are future work.

To conclude, we developed and evaluated a CNN-based-approach to fully automatically segment catheters in live fluoroscopic images. With execution times within 75 ms, this method is promising for use in real-time catheter detection.

Conflict of Interest. The authors declare that they have no conflict of interest.

References

1. Ambrosini, P., Smal, I., Ruijters, D., Niessen, W.J., Moelker, A., van Walsum, T.: A hidden Markov model for 3D catheter tip tracking with 2D X-ray catheterization sequence and 3D rotational angiography. *IEEE Trans. Med. Imag.* **36**(3), 757–768 (2017)
2. Baert, S.A., Viergever, M.A., Niessen, W.J.: Guide-wire tracking during endovascular interventions. *IEEE Trans. Med. Imag.* **22**(8), 965–972 (2003)
3. Baur, C., Albarqouni, S., Demirci, S., Navab, N., Fallavollita, P.: CathNets: detection and single-view depth prediction of catheter electrodes. In: Zheng, G., Liao, H., Jannin, P., Cattin, P., Lee, S.-L. (eds.) *MIAR 2016*. LNCS, vol. 9805, pp. 38–49. Springer, Cham (2016). doi:[10.1007/978-3-319-43775-0_4](https://doi.org/10.1007/978-3-319-43775-0_4)
4. Chang, P.L., Rolls, A., De Praetere, H., Vander Poorten, E., Riga, C.V., Bicknell, C.D., Stoyanov, D.: Robust catheter and guidewire tracking using b-spline tube model and pixel-wise posteriors. *IEEE Robot. Autom. Lett.* **1**(1), 303–308 (2016)
5. Chen, B.J., Wu, Z., Sun, S., Zhang, D., Chen, T.: Guidewire tracking using a novel sequential segment optimization method in interventional X-ray videos. In: *Proceedings of IEEE International Symposium Biomedical Imaging*, pp. 103–106 (2016)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of IEEE Computer Society Conference on computer vision and pattern recognition*, pp. 770–778 (2016)
7. Heibel, H., Glocker, B., Groher, M., Pfister, M., Navab, N.: Interventional tool tracking using discrete optimization. *IEEE Trans. Med. Imag.* **32**(3), 544–555 (2013)
8. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *Proceedings of ICML*, vol. 37, pp. 448–456 (2015)
9. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: *Proceedings of IEEE Conference on CVPR*, pp. 3431–3440 (2015)
10. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: *Proceedings of International Conference on 3D Vision*, pp. 565–571. IEEE (2016)
11. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015*. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). doi:[10.1007/978-3-319-24574-4_28](https://doi.org/10.1007/978-3-319-24574-4_28)
12. Ruijters, D., Homan, R., Mielekamp, P., Van de Haar, P., Babic, D.: Validation of 3D multimodality roadmapping in interventional neuroradiology. *Phys. Med. Biol.* **56**(16), 5335 (2011)

13. Slabaugh, G., Kong, K., Unal, G., Fang, T.: Variational guidewire tracking using phase congruency. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007. LNCS, vol. 4792, pp. 612–619. Springer, Heidelberg (2007). doi:[10.1007/978-3-540-75759-7_74](https://doi.org/10.1007/978-3-540-75759-7_74)
14. Springenberg, J.T., Dosovitskiy, A., Brox, T., Riedmiller, M.: Striving for simplicity: the all convolutional net. arXiv preprint [arXiv:1412.6806](https://arxiv.org/abs/1412.6806) (2014)
15. Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**(1), 1929–1958 (2014)
16. Wagner, M.G., Strother, C.M., Mistretta, C.A.: Guidewire path tracking and segmentation in 2D fluoroscopic time series using device paths from previous frames. In: Proceedings of SPIE, p. 97842B (2016)
17. Wang, P., Chen, T., Zhu, Y., Zhang, W., Zhou, S., Comaniciu, D.: Robust guidewire tracking in fluoroscopy. In: Proceedings of IEEE Computer Society CVPR, pp. 691–698 (2009)
18. Wu, X., Housden, J., Ma, Y., Razavi, B., Rhode, K., Rueckert, D.: Fast catheter segmentation from echocardiographic sequences based on segmentation from corresponding X-ray fluoroscopy for cardiac catheterization interventions. *IEEE Trans. Med. Imag.* **34**(4), 861–876 (2015)
19. Zhang, T., Suen, C.Y.: A fast parallel algorithm for thinning digital patterns. *Commun. ACM* **27**(3), 236–239 (1984)